

# Bounds on the Constant in the Mean Central Limit Theorem

Larry Goldstein  
University of Southern California

June 28, 2009

## Abstract

Let  $X_1, \dots, X_n$  be independent with mean zero, finite variances  $\sigma_1^2, \dots, \sigma_n^2$  and finite absolute third moments,  $F_n$  the distribution function of  $(X_1 + \dots + X_n)/\sigma$  where  $\sigma^2 = \sum_{i=1}^n \sigma_i^2$ , and  $\Phi$  that of the standard normal. Then the  $L^1$  distance between  $F_n$  and  $\Phi$  satisfies

$$\|F_n - \Phi\|_1 \leq \frac{1}{\sigma^3} \sum_{i=1}^n E|X_i|^3.$$

In particular, when  $X_1, \dots, X_n$  are identically distributed with variance  $\sigma^2$ ,

$$\|F_n - \Phi\|_1 \leq \frac{E|X_1|^3}{\sigma^3 \sqrt{n}} \quad \text{for all } n \in \mathbb{N},$$

corresponding to an  $L^1$  Berry Esseen constant of 1. A lower bound of

$$\frac{2\sqrt{\pi}(2\Phi(1) - 1) - (\sqrt{\pi} + \sqrt{2}) + 2e^{-1/2}\sqrt{2}}{\sqrt{\pi}} = 0.535377\dots$$

on the smallest possible constant is provided.

## 1 Introduction

The classical central limit allows the approximation of the distribution of sums of ‘comparable’ independent real valued random variables by the normal. As this theorem is an asymptotic, it provides no information as to whether the resulting approximation is useful. For that purpose one may turn to the Berry-Esseen theorem, the most classical version giving supremum norm bounds between the distribution function of the normalized sum and that of the standard normal. Various authors have also considered Berry-Esseen type bounds using other metrics, and in particular bounds in  $L^p$ . The case  $p = 1$ , where the value

$$\|F - G\|_1 = \int_{-\infty}^{\infty} |F(x) - G(x)| dx$$

---

<sup>0</sup>AMS 2000 subject classifications: Primary 60F05, 60F25.

<sup>0</sup>Key words and phrases: Stein’s method, Berry Esseen constant.

is used to measure the distance between distribution functions  $F$  and  $G$ , is of some particular interest, and results using this metric are known as mean central limit theorems, see, for instance, [12], [4], [11] and [1]; the latter three of these works consider nonindependent summand variables. One motivation for studying  $L^1$  bounds is that combined with one of type  $L^\infty$ , bounds on  $L^p$  distance for all  $p \in (1, \infty)$  may be obtained by the inequality

$$\|F - G\|_p^p \leq \|F - G\|_\infty^{p-1} \|F - G\|_1.$$

For  $\sigma \in (0, \infty)$  let  $\mathcal{F}_\sigma$  be the collection of distributions with mean zero, variance  $\sigma^2$ , and finite absolute third moment. We prove the following Berry Esseen type result for the mean central limit theorem.

**Theorem 1.1** *For  $n \in \mathbb{N}$  let  $X_1, \dots, X_n$  be independent mean zero random variables with distributions  $G_1 \in \mathcal{F}_{\sigma_1}, \dots, G_n \in \mathcal{F}_{\sigma_n}$ , and let  $F_n$  be the distribution of*

$$W = \frac{1}{\sigma} \sum_{i=1}^n X_i \quad \text{where} \quad \sigma^2 = \sum_{i=1}^n \sigma_i^2.$$

*Then*

$$\|F_n - \Phi\|_1 \leq \frac{1}{\sigma^3} \sum_{i=1}^n E|X_i|^3.$$

*In particular, when  $X_1, \dots, X_n$  are identically distributed with distribution  $G \in \mathcal{F}_1$ ,*

$$\|F_n - \Phi\|_1 \leq \frac{E|X_1|^3}{\sqrt{n}} \quad \text{for all } n \in \mathbb{N}.$$

For the case where all variables are identically distributed as  $X$  having distribution  $G$ , letting

$$c_m = \inf \left\{ C : \frac{\sqrt{n}\sigma^3 \|F_n - \Phi\|_1}{E|X|^3} \leq C \quad \text{for all } G \in \mathcal{F}_1 \text{ and } n \geq m \right\}, \quad (1)$$

the second part of Theorem 1.1 can be restated as the upper bound  $c_1 \leq 1$ . We also provide the following lower bound

**Theorem 1.2** *With  $c_1$  given by (1) for  $m = 1$ ,*

$$c_1 \geq \frac{2\sqrt{\pi}(2\Phi(1) - 1) - (\sqrt{\pi} + \sqrt{2}) + 2e^{-1/2}\sqrt{2}}{\sqrt{\pi}} = 0.535377\dots \quad (2)$$

Clearly the elements of the sequence  $\{c_m\}_{m \geq 1}$  are nonnegative and decreasing in  $m$ , and therefore has a limit, say  $c_\infty$ . Regarding limiting behavior Esseen [3] showed that

$$\lim_{n \rightarrow \infty} n^{1/2} \|F_n - \Phi\|_1 = A(G)$$

for an explicit constant  $A(G)$  depending only on  $G$ . Zolotarev [18] provides the representation

$$A(G) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-1/2}^{1/2} \int_{-\infty}^{\infty} \left| \frac{\omega}{2}(1 - x^2) + hu \right| e^{-x^2/2} dx du \quad (3)$$

where  $\omega = |EX^3|/(3\sigma^2)$  and  $h$  is the span of the distribution  $G$  in case  $G$  is lattice, and zero otherwise. Zolotarev obtains

$$\sup_{G \in \mathcal{F}_\sigma} \frac{\sigma^3 A(G)}{E|X|^3} = \frac{1}{2},$$

showing  $c_\infty = 1/2$ , giving the asymptotic  $L^1$  Berry Esseen constant value.

Here the focus is on nonasymptotic constants, and in particular on the constant  $c_1$  which gives a bound for all  $n \in \mathbb{N}$ . Theorem 1.1 is shown using Stein's method (see [15], [17]) which uses the characterizing equation (5) for the normal, and an associated differential equation to obtain bounds on the normal approximation. More particularly, we employ the zero bias transformation, introduced in [9], and the evaluation of a Stein functional, as in [13]. In [9] it was shown that for all  $X$  with mean zero and finite non-zero variance  $\sigma^2$  there exists a unique distribution for a random variable  $X^*$  such that

$$\sigma^2 E f'(X^*) = E[X f(X)] \quad (4)$$

for all absolutely continuous functions  $f$  for which these expectations exist. The zero bias transformation, mapping the distribution of  $X$  to that of  $X^*$ , was motivated by the Stein characterization of the normal distribution [16], which states that  $Z$  is normal with mean zero and variance  $\sigma^2$  if and only if

$$\sigma^2 E f'(Z) = E[Z f(Z)] \quad (5)$$

for all absolutely continuous functions  $f$  for which these expectations exist. Hence, the mean zero normal with variance  $\sigma^2$  is the unique fixed point of the zero bias transformation. How closeness to normality may be measured by the closeness of a distribution to its transform, and applications, are the topics of [5] and [6].

As shown in [9] and [7], for a random variable  $X$  with  $EX = 0$  and  $\text{Var}(X) = \sigma^2$ , the distribution of  $X^*$  is absolutely continuous with density and distribution functions given by, respectively,

$$g^*(x) = \sigma^{-2} E[X \mathbf{1}(X > x)] \quad \text{and} \quad G^*(x) = \sigma^{-2} E[X(X - x) \mathbf{1}(X \leq x)]. \quad (6)$$

Theorem 1.1 results by showing that the functional

$$B(G) = \frac{2\sigma^2 \|G^* - G\|_1}{E|X|^3} \quad (7)$$

is bounded by 1 for all  $X$  with distribution  $G \in \mathcal{F}_\sigma$ . As in (3) one may write out a more 'explicit' form for  $B(G)$  using (6) and expressions for the moments on which  $B(G)$  depends, however such expressions appear to be of little value for the purposes of proving Theorem 1.1. In turn, the proof here employs convexity properties of  $B(G)$  which depend on the behavior of the zero bias transformation on mixtures. We note also that the functional  $B(G)$  is somewhat different from  $A(G)$ ; for instance,  $A(G)$  is zero for all nonlattice distributions with vanishing third moment, whereas  $B(G)$  is zero only for mean zero normal distributions. Parallels to the current work appear in [13] where a different type of Stein functional was studied using somewhat similar methods, see in particular Proposition 4.1 there.

Let  $\mathcal{L}(X)$  denote the distribution of a random variable  $X$ . Since the  $L^1$  distance scales, that is, since for all  $a \in \mathbb{R}$

$$\|\mathcal{L}(aX) - \mathcal{L}(aY)\|_1 = |a| \|\mathcal{L}(X) - \mathcal{L}(Y)\|_1, \quad (8)$$

by replacing  $\sigma_i^2$  by  $\sigma_i^2/\sigma^2$  and  $\|G_i^* - G_i\|_1$  by  $\|G_i^* - G_i\|_1/\sigma$  in equation (16) of Theorem 2.1 of [7] we obtain

**Proposition 1.1** *Under the hypotheses of Theorem 1.1,*

$$\|F_n - \Phi\|_1 \leq \frac{1}{\sigma^3} \sum_{i=1}^n B(G_i) E|X_i|^3.$$

For  $\mathcal{F}$  a collection of mean zero distributions with finite absolute third moments let

$$B(\mathcal{F}) = \sup_{G \in \mathcal{F}} B(G).$$

Clearly, Theorem 1.1 follows immediately from Proposition 1.1 and the following result.

**Lemma 1.1** *For all  $\sigma \in (0, \infty)$ ,*

$$B(\mathcal{F}_\sigma) = 1.$$

The equality to 1 in Lemma 1.1 improves the upper bound of 3 shown in [7]. Though our interest here is in best universal constants, we note that Proposition 1.1 provides  $B(G)$  as a distribution specific  $L^1$  Berry-Esseen constant in that

$$\|F_n - \Phi\|_1 \leq \frac{B(G) E|X_1|^3}{\sigma^3 \sqrt{n}} \quad \text{for all } n \in \mathbb{N}$$

when  $X_1, \dots, X_n$  are identically distributed according to  $G \in \mathcal{F}_\sigma$ . For instance  $B(G) = 1/3$  when  $G$  is a mean zero uniform distribution, and  $B(G) = 1$  when  $G$  is a mean zero two point distribution, see Corollary 2.1 of [7], and Lemmas 1.2 and 1.3 below.

We close this section with two preliminaries. The first collects some facts shown in [7], and the second demonstrates that to prove Lemma 1.1 it suffices to consider the class of random variables  $\mathcal{F}_1$ . Then, following Hoeffding [10] (see also [13]) in Section 2 we use a continuity property of  $B(G)$  to show that its supremum over  $\mathcal{F}_1$  is attained on finitely supported distributions. Exploiting a convexity type property of the zero bias transformation on mixtures over distributions having equal variances we reduce the calculation further to the calculation of the supremum over  $D_3$ , the collection of all mean zero distributions with variance 1, supported on at most three points. As three point distributions are in general a mixture of two two point distributions with unequal variances, an additional argument is given in Section 3 where a coupling of an  $X$  with distribution  $G \in D_3$  to a variable  $X^*$  having the  $X$  zero bias distribution is constructed, using the optimal  $L^1$  couplings on the component two point distributions of which  $G$  is the mixture, in order to obtain  $B(G) \leq 1$  for all  $G \in D_3$ . Theorem 1.2, the lower bound on  $c_1$ , is calculated in Section 4.

The following simple formula will be of some use. For  $l, a$  and  $b$  nonnegative,

$$\int_0^l \left| (a+b) \frac{u}{l} - a \right| du = \frac{l}{2} \frac{a^2 + b^2}{a+b}. \quad (9)$$

**Lemma 1.2** *Let  $G$  be the distribution of a nontrivial mean zero random variable  $X$  supported on the two points  $x < y$ . Then  $X^*$  is uniformly distributed on  $[x, y]$ ,*

$$EX^2 = -xy, \quad E|X^3| = \frac{-xy(y^2 + x^2)}{y - x}, \quad \text{and} \quad \|\mathcal{L}(X^*) - \mathcal{L}(X)\|_1 = \frac{1}{2} \frac{y^2 + x^2}{y - x}.$$

In particular  $B(G) = 1$  and

$$B(\mathcal{F}_1) \geq 1.$$

**Proof:** Being nontrivial  $G$  has positive variance, and from (6) we see that the density  $g^*$  of  $G^*$ , which is proportional to  $E[X\mathbf{1}(X > x)]$ , is zero outside  $[x, y]$  and constant within it, so  $G^*(w) = (w - x)/(y - x)$  for  $w \in [x, y]$ . That  $G$  has mean zero implies that the support points  $x$  and  $y$  satisfy  $x < 0 < y$  and that  $G$  gives positive probability  $y/(y - x)$  and  $-x/(y - x)$  to  $x$  and  $y$  respectively. The moment identities are immediate.

Making the change of variable  $u = w - x$  and applying (9) with  $a = y/(y - x)$ ,  $b = -x/(y - x)$  and  $l = y - x$  yields

$$\|\mathcal{L}(X^*) - \mathcal{L}(X)\|_1 = \int_x^y \left| \frac{w - x}{y - x} - \frac{y}{y - x} \right| dw = \frac{1}{2} \left( \frac{y^2 + x^2}{y - x} \right),$$

and (7) now gives  $B(G) = 1$ . ■

**Lemma 1.3** *Let  $G \in \mathcal{F}_\sigma$  for some  $\sigma \in (0, \infty)$ , let  $X$  have distribution  $G$ , and for  $a \neq 0$  let  $G_a$  denote the distribution of  $aX$ . Then  $B(G_a) = B(G)$  and in particular*

$$B(\mathcal{F}_\sigma) = B(\mathcal{F}_1) \quad \text{for all } \sigma \in (0, \infty).$$

**Proof:** That  $aX^*$  has the same distribution as  $(aX)^*$  follows from (4). Now the identities  $\sigma_{aX}^2 = a^2\sigma_X^2$ ,  $E|aX|^3 = |a|^3E|X|^3$  and (8) imply the first claim. Since

$$\{B(G) : G \in \mathcal{F}_\sigma\} = \{B(G) : G \in \mathcal{F}_1\},$$

taking supremum completes the proof. ■

## 2 Reduction to three point distributions

Let  $(S, \Sigma)$  be a measurable space, and let  $\{m_s\}_{s \in S}$  be a collection of probability measures on  $\mathbb{R}$  such that for each Borel subset  $A \subset \mathbb{R}$  the function from  $S$  to  $[0, 1]$  given by

$$s \rightarrow m_s(A)$$

is measurable. When  $\mu$  is a probability measure on  $(S, \Sigma)$ , the set function given by

$$m_\mu(A) = \int_S m_s(A) \mu(ds)$$

is a probability measure, and called the  $\mu$  mixture of  $\{m_s\}_{s \in S}$ . With some slight abuse of notation, we let  $E_\mu$  and  $E_s$  denote expectations with respect to  $m_\mu$  and  $m_s$  and let  $X_\mu$  and

$X_s$  be random variables with distributions  $m_\mu$  and  $m_s$ , respectively. For instance, for all functions  $f$  which are integrable with respect to  $\mu$  we have

$$E_\mu f(X) = \int E_s f(X) \mu(ds) \quad \text{which we also write as} \quad Ef(X_\mu) = \int Ef(X_s) \mu(ds).$$

In particular, if  $\{m_s\}_{s \in S}$  is a collection of mean zero distributions with variances  $\sigma_s^2 = EX_s^2$  and absolute third moments  $\gamma_s = E|X_s^3|$ , the mixture distribution  $m_\mu$  has variance  $\sigma_\mu^2$  and third absolute moment  $\gamma_\mu$  given by

$$\sigma_\mu^2 = \int_S \sigma_s^2 d\mu \quad \text{and} \quad \gamma_\mu = \int_S \gamma_s d\mu,$$

where both may be infinite. Note that  $\sigma_\mu^2 < \infty$  implies  $\sigma_s^2 < \infty$   $\mu$ -almost surely, and therefore that  $m_s^*$ , the  $m_s$  zero bias distribution, exists  $\mu$ -almost surely.

Theorem 2.1 shows that the zero bias distribution of a mixture is a mixture of zero bias distributions with mixing measure the original measure weighted by the variance and rescaled. Define (arbitrarily, see Remark 2.1) the zero bias distribution of  $\delta_0$ , a point mass at zero, to be  $\delta_0$ . Write  $X =_d Y$  when  $X$  and  $Y$  have the same distribution.

**Theorem 2.1** *Let  $\{m_s, s \in S\}$  be a collection of mean zero distributions on  $\mathbb{R}$  and  $\mu$  a probability measure on  $S$  such that the variance  $\sigma_\mu^2$  of the mixture distribution is positive and finite. Then  $m_\mu^*$ , the  $m_\mu$  zero bias distribution exists and is given by the mixture*

$$m_\mu^* = \int m_s^* d\nu \quad \text{where} \quad \frac{d\nu}{d\mu} = \frac{\sigma_s^2}{\sigma_\mu^2}.$$

*In particular,  $\nu = \mu$  if and only if  $\sigma_s^2$  is a constant  $\mu$  a.s.*

**Proof:** The distribution  $m_\mu^*$  exists as  $m_\mu$  has mean zero and finite nonzero variance. Let  $X_\mu^*$  have the  $m_\mu$  zero bias distribution, and let  $Y$  have distribution  $m_\mu^*$ . For any absolutely continuous function  $f$  for which the expectations below exist,

$$\begin{aligned} \sigma_\mu^2 E f'(X_\mu^*) &= E X_\mu f(X_\mu) \\ &= \int E X_s f(X_s) d\mu \\ &= \int \sigma_s^2 E f'(X_s^*) d\mu \\ &= \sigma_\mu^2 \int E f'(X_s^*) d\nu \\ &= \sigma_\mu^2 E f'(Y). \end{aligned}$$

Since  $E f'(X_\mu^*) = E f'(Y)$  for all such  $f$  we conclude  $X_\mu^* =_d Y$ . ■

**Remark 2.1** *If  $m_s = \delta_0$  for any  $s \in S$  then  $\sigma_s^2 = 0$ , and therefore*

$$\nu\{s \in S : m_s = \delta_0\} = 0.$$

*Hence the mixture  $X_\mu^*$  gives zero weight to  $m_s^*$  for all such  $s$ , showing that  $(\delta_0)^*$  may be defined arbitrarily.*

We now recall an equivalent form of the  $L^1$  distance involving expectations of Lipschitz functions  $L$  on  $\mathbb{R}$ ,

$$\|F - G\|_1 = \sup_{f \in L} |Ef(X) - Ef(Y)| \quad \text{where} \quad L = \{f : |f(x) - f(y)| \leq |x - y|\}, \quad (10)$$

and  $X$  and  $Y$  have distribution  $F$  and  $G$ , respectively. With a slight abuses of notation we may write  $B(X)$  in place of  $B(G)$  when  $X$  has distribution  $G$ .

**Theorem 2.2** *Let  $X_\mu$  be the  $\mu$  mixture of a collection  $\{X_s, s \in S\}$  of mean zero, variance 1 random variables satisfying  $E|X_\mu^3| < \infty$ . Then*

$$B(X_\mu) \leq \sup_{s \in S} B(X_s). \quad (11)$$

*In particular, if  $\mathcal{C}$  is a collection of mean zero, variance 1 random variables with finite absolute third moments and  $\mathcal{D} \subset \mathcal{C}$  such that every distribution in  $\mathcal{C}$  can be represented as a mixture of distributions in  $\mathcal{D}$ , then*

$$B(\mathcal{C}) = B(\mathcal{D}). \quad (12)$$

**Proof:** Since the variances  $\sigma_s^2$  of  $X_s$  are constant the distribution  $X_\mu^*$  is the  $\mu$  mixture of  $\{X_s^*, s \in S\}$  by Theorem 2.1. Hence, applying (10),

$$\begin{aligned} \|\mathcal{L}(X_\mu^*) - \mathcal{L}(X_\mu)\|_1 &= \sup_{f \in L} |Ef(X_\mu^*) - Ef(X_\mu)| \\ &= \sup_{f \in L} \left| \int_S Ef(X_s^*) d\mu - \int_S Ef(X_s) d\mu \right| \\ &\leq \sup_{f \in L} \int_S |Ef(X_s^*) - Ef(X_s)| d\mu \\ &\leq \sup_{f \in L} \int_S \|\mathcal{L}(X_s^*) - \mathcal{L}(X_s)\|_1 d\mu \\ &= \int_S \|\mathcal{L}(X_s^*) - \mathcal{L}(X_s)\|_1 d\mu. \end{aligned} \quad (13)$$

Now let  $\tau$  be the measure on  $(S, \Sigma)$  which is absolutely continuous with respect to  $\mu$  with Radon Nikodym derivative

$$\frac{d\tau}{d\mu} = \frac{E|X_s^3|}{E|X_\mu^3|}. \quad (14)$$

This relation defines a probability measure as  $E|X_\mu^3| = \int_S E|X_s^3| d\mu$ . Noting also that  $\text{Var}(X_\mu) = \int_S EX_s^2 d\mu = 1$ , applying (13) we find

$$\begin{aligned} B(X_\mu) &= \frac{2\|\mathcal{L}(X_\mu^*) - \mathcal{L}(X_\mu)\|_1}{E|X_\mu^3|} \\ &\leq \frac{\int_S 2\|\mathcal{L}(X_s^*) - \mathcal{L}(X_s)\|_1 d\mu}{E|X_\mu^3|} \\ &= \frac{\int_S B(X_s) E|X_s^3| d\mu}{E|X_\mu^3|} \\ &= \int_S B(X_s) d\tau \\ &\leq \sup_{s \in S} B(X_s), \end{aligned} \quad (15)$$

proving (11).

Regarding (12), clearly  $B(\mathcal{D}) \leq B(\mathcal{C})$ , and the reverse inequality follows from (11).  $\blacksquare$

**Remark 2.2** *The supremum over  $S$  in (15), and therefore in the theorem, can be replaced with essential supremum, with respect to  $\tau$  in (14), over  $S$ .*

Note that no bound of the type provided by Theorem 2.2 holds in general when taking mixtures of variables that have unequal variances. In particular, if  $X_s \sim \mathcal{N}(0, \sigma_s^2)$  and  $\sigma_s^2$  is not constant in  $s$ , then  $X_\mu$  is a mixture of normals with unequal variances, which is not normal. Hence, in this case  $B(X_\mu) > 0$ , whereas  $B(X_s) = 0$  for all  $s$ .

To apply Theorem 2.2 to reduce the computation of  $B(\mathcal{F}_1)$  to finitely supported distributions we apply the following continuity property of the zero bias transformation, see Lemma 5.2 in [8]. We write  $X_n \Rightarrow X$  for the convergence of  $X_n$  to  $X$  in distribution.

**Lemma 2.1** *Let  $X$  and  $X_n, n = 1, 2, \dots$  be mean zero random variables with finite, nonzero variances. If*

$$X_n \Rightarrow_d X \quad \text{and} \quad \lim_{n \rightarrow \infty} EX_n^2 = EX^2,$$

*then*

$$X_n^* \Rightarrow_d X^*.$$

For a distribution function  $F$  let

$$F^{-1}(w) = \sup\{a : F(a) < w\} \quad \text{for all } w \in (0, 1). \quad (16)$$

If  $U$  is uniform on  $[0, 1]$  then  $F^{-1}(U)$  has distribution function  $F$ , and if  $X_n$  and  $X$  have distribution functions  $F_n$  and  $F$  respectively and  $X_n \Rightarrow X$  then  $F_n^{-1}(U) \rightarrow F^{-1}(U)$  a.s (see, e.g., Theorem 2.1 of [2]). For distribution functions  $F$  and  $G$ ,

$$\|F - G\|_1 = \inf E|X - Y| \quad (17)$$

where the infimum is over all joint distributions on  $X, Y$  which have marginals  $F$  and  $G$  respectively, and the variables  $F^{-1}(U)$  and  $G^{-1}(U)$  achieve the minimal  $L^1$  coupling, that is,

$$\|F - G\|_1 = E|F^{-1}(U) - G^{-1}(U)|, \quad (18)$$

see [14] for details.

With the use of Lemma 2.1 we are able to prove the following continuity property of the functional  $B(X)$ .

**Lemma 2.2** *Let  $X$  and  $X_n, n \in \mathbb{N}$  be mean zero random variables with finite, nonzero absolute third moments. If*

$$X_n \Rightarrow_d X, \quad \lim_{n \rightarrow \infty} EX_n^2 = EX^2 \quad \text{and} \quad E|X_n^3| \rightarrow E|X^3| \quad (19)$$

*then*

$$B(X_n) \rightarrow B(X) \quad \text{as } n \rightarrow \infty.$$

**Proof:** By Lemma 2.1 we have  $X_n^* \Rightarrow X^*$ . Let  $U$  be a uniformly distributed variable and set

$$(Y, Y_n, Y^*, Y_n^*) = (F_X^{-1}(U), F_{X_n}^{-1}(U), F_{X^*}^{-1}(U), F_{X_n^*}^{-1}(U))$$

where  $F_X$  denotes the distribution function of  $F_X$ , and so forth. Then  $(Y, Y_n, Y^*, Y_n^*) =_d (X, X_n, X^*, X_n^*)$ ,  $Y_n \rightarrow_{a.s.} Y$  and  $Y_n^* \rightarrow_{a.s.} Y^*$  and by (18)

$$\|\mathcal{L}(X_n) - \mathcal{L}(X)\|_1 = E|Y_n^* - Y_n| \quad \text{and} \quad \|\mathcal{L}(X^*) - \mathcal{L}(X)\| = E|Y^* - Y|.$$

By (4) with  $f(x) = x^2 \text{sgn}(x)$  we find, for  $Y$  for example, that

$$E|Y^3| = 2\text{Var}(Y)E|Y^*|.$$

Hence  $EY_n^2 = EX_n^2 \rightarrow EX^2 = EY^2$

$$E|Y_n^*| = \frac{E|Y_n^3|}{2EY_n^2} = \frac{E|X_n^3|}{2EX_n^2} \rightarrow \frac{E|X^3|}{2EX^2} = \frac{E|Y^3|}{2EY^2} = E|Y^*| \quad \text{as } n \rightarrow \infty.$$

Hence  $\{Y_n\}_{n \in \mathbb{N}}$  and  $\{Y_n^*\}_{n \in \mathbb{N}}$  are uniformly integrable, so  $\{Y_n^* - Y_n\}_{n \in \mathbb{N}}$  is uniformly integrable. As  $Y_n^* - Y_n \rightarrow_{a.s.} Y^* - Y$  as  $n \rightarrow \infty$ ,

$$\lim_{n \rightarrow \infty} \|\mathcal{L}(X_n) - \mathcal{L}(X)\|_1 = \lim_{n \rightarrow \infty} E|Y_n^* - Y_n| = E|Y^* - Y| = \|\mathcal{L}(X^*) - \mathcal{L}(X)\|. \quad (20)$$

Combining (20) with the convergence of the variances and the absolute third moments as provided by (19) the proof is complete.  $\blacksquare$

Lemmas 2.3 and 2.4 borrow much from Theorem 2.1 of [10], the latter lemma indeed being implicit. The results of [10] are not applied directly as  $B(G)$  is not expressed as the expectation of  $K(X)$  for some  $K$  when  $\mathcal{L}(X) = G$ . For  $m \geq 2$  let  $D_m$  denote the collection of all mean zero, variance 1 distributions which are supported on at most  $m$  points.

### Lemma 2.3

$$B(\mathcal{F}_1) = B\left(\bigcup_{m \geq 3} D_m\right).$$

**Proof:** Letting  $\mathcal{M}$  be the collection of distributions in  $\mathcal{F}_1$  which have compact support we first show that

$$B(\mathcal{F}_1) \leq B(\mathcal{M}). \quad (21)$$

Let  $\mathcal{L}(X) \in \mathcal{F}_1$  be given and for  $n \in \mathbb{N}$  set  $Y_n = X\mathbf{1}_{|X| \leq n}$ . Clearly  $Y_n \Rightarrow_d X$ . As  $E|X^3| < \infty$  and  $|Y_n^p| \leq |X^p|$  for all  $p \geq 0$ , by the dominated convergence theorem

$$EY_n \rightarrow EX = 0, \quad EY_n^2 \rightarrow EX^2 = 1 \quad \text{and} \quad E|Y_n^3| \rightarrow E|X^3| \quad \text{as } n \rightarrow \infty. \quad (22)$$

Letting

$$X_n = Y_n - EY_n \quad (23)$$

we have  $X_n \Rightarrow X$  by Slutsky's theorem, so, in view of (22) the hypotheses of Lemma 2.2 are satisfied, yielding

$$B(X_n) \rightarrow B(X) \quad \text{as } n \rightarrow \infty, \quad \text{with } \{X_n\}_{n \in \mathbb{N}} \subset \mathcal{M},$$

showing (21).

Now consider  $\mathcal{L}(X) \in \mathcal{M}$ , so that  $|X| \leq M$  a.s. for some  $M > 0$ . For each  $n \in \mathbb{N}$  let

$$Y_n = \sum_{k \in \mathbb{Z}} \frac{k}{2^n} \mathbf{1}\left(\frac{k-1}{2^n} < X \leq \frac{k}{2^n}\right).$$

Since  $|X| \leq M$  a.s., each  $Y_n$  is supported on finitely many points and  $|Y_n| \leq 2M$  for all  $n$  sufficiently large. Clearly  $Y_n \rightarrow X$  a.s. and (22) holds by the bounded convergence theorem. Now defining  $X_n$  by (23) the hypotheses of Lemma 2.2 are satisfied, yielding

$$B(X_n) \rightarrow B(X) \quad \text{as } n \rightarrow \infty, \text{ with } \{X_n\}_{n \in \mathbb{N}} \subset \bigcup_{m \geq 3} D_m.$$

showing  $B(\mathcal{M}) \leq B(\bigcup_{m \geq 3} D_m)$ . Combining this inequality with (21) yields  $B(\mathcal{F}_1) \leq B(\bigcup_{m \geq 3} D_m)$  and therefore the lemma, the reverse inequality being obvious. ■

**Lemma 2.4** *Every distribution in  $\bigcup_{m \geq 3} D_m$  can be expressed as a finite mixture of  $D_3$  distributions.*

**Proof:** The lemma is trivially true for  $m = 3$  so consider  $m > 3$  and assume that the lemma holds for all integers from 3 to  $m - 1$ .

The distribution of any  $X \in D_m$  is determined by the supporting values  $a_1 < \dots < a_m$  and a vector of probabilities  $\mathbf{p} = (p_1, \dots, p_m)'$ . If any of the components of  $\mathbf{p}$  are zero then  $X \in D_k$  for  $k < m$  and the induction would be finished, so assume all components of  $\mathbf{p}$  are strictly positive. As  $X \in D_m$  the vector  $\mathbf{p}$  must satisfy

$$A\mathbf{p} = \mathbf{c} \quad \text{where} \quad A = \begin{bmatrix} 1 & 1 & \dots & 1 \\ a_1 & a_2 & \dots & a_m \\ a_1^2 & a_2^2 & \dots & a_m^2 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Since  $A \in \mathbb{R}^{3 \times m}$  with  $m > 3$ ,  $\mathcal{N}(A) \neq \{0\}$ , that is, there exists  $\mathbf{v} \neq 0$  with

$$A\mathbf{v} = 0. \tag{24}$$

Since  $\mathbf{v} \neq 0$  and the equation specified by the first row of  $A$  is exactly that  $\sum_i v_i = 0$ , the vector  $\mathbf{v}$  contains both positive and negative numbers. Since the vector  $\mathbf{p}$  has strictly positive components, the numbers  $t_1$  and  $t_2$  given by

$$t_1 = \inf\{t > 0 : \min_i (p_i + tv_i) \geq 0\} \quad \text{and} \quad t_2 = \inf\{t > 0 : \min_i (p_i - tv_i) \geq 0\}$$

are both strictly positive. Note that

$$\mathbf{p}_1 = \mathbf{p} + t_1 \mathbf{v} \quad \text{and} \quad \mathbf{p}_2 = \mathbf{p} - t_2 \mathbf{v}$$

satisfy

$$A\mathbf{p}_i = A(\mathbf{p} + t_i \mathbf{v}) = A\mathbf{p} = \mathbf{c} \quad \text{for } i \in \{1, 2\}$$

by (24), so that  $\mathbf{p}_1$  and  $\mathbf{p}_2$  are probability vectors, as their components are nonnegative and sum to one. Additionally, the the corresponding distribution have mean zero and variance

1, and in each of these two vectors at least one component has been set to zero. Hence we may express the  $m$  point probability vector  $\mathbf{p}$  as the mixture

$$\mathbf{p} = \frac{t_2}{t_1 + t_2} \mathbf{p}_1 + \frac{t_1}{t_1 + t_2} \mathbf{p}_2$$

of probability vectors on at most  $m - 1$  support points, thus showing  $X$  to be the mixture of two distributions in  $D_{m-1}$ , completing the induction.  $\blacksquare$

The following theorem is an immediate consequence of Theorem 2.2 and Lemmas 2.3 and 2.4.

**Theorem 2.3**

$$B(\mathcal{F}_1) = B(D_3).$$

Hence we now restrict attention to  $D_3$ .

### 3 Bound for $D_3$ distributions

Clearly, Lemma 1.1 follows from Lemma 1.3, Theorem 2.3 and Theorem 3.1 below, which shows  $B(D_3) = 1$ . We prove Theorem 3.1 with the help of the following result.

**Lemma 3.1** *Let  $x < y < 0 < z$  and let  $m_1$  and  $m_0$  be the unique mean zero distributions with support  $\{x, z\}$  and  $\{y, z\}$  respectively, that is,*

$$m_1(\{w\}) = \begin{cases} \frac{z}{z-x} & \text{if } w = x \\ \frac{-x}{z-x} & \text{if } w = z \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad m_0(\{w\}) = \begin{cases} \frac{z}{z-y} & \text{if } w = y \\ \frac{-y}{z-y} & \text{if } w = z \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\|m_1^* - m_0\|_1 \leq \|m_1^* - m_1\|_1. \quad (25)$$

**Proof:** By Lemma 1.2

$$\begin{aligned} \|m_1^* - m_1\|_1 &= \frac{z^2 + x^2}{2(z-x)} = \frac{(z^2 + x^2)(z-y)^2}{2(z-x)(z-y)^2} = \frac{z^4 - 2yz^3 + y^2z^2 + x^2z^2 - 2x^2yz + x^2y^2}{2(z-x)(z-y)^2} \\ &= \frac{(z^4 - 2yz^3 + x^2z^2 - 2x^2yz) + y^2z^2 + x^2y^2}{2(z-x)(z-y)^2}. \end{aligned} \quad (26)$$

Let  $F_1, F_0, F_1^*$  and  $F_0^*$  denote the distribution functions of  $m_1, m_0, m_1^*$  and  $m_0^*$  respectively. By Lemma 1.2  $m_1^*$  and  $m_0^*$  are uniform over  $[x, z]$  and  $[y, z]$ , respectively. Letting  $J_1 = [x, y]$  and  $J_2 = [y, z]$  we have

$$\|m_1^* - m_0\|_1 = I_1 + I_2 \quad \text{where} \quad I_i = \int_{J_i} |F_1^*(w) - F_0(w)| dw \quad \text{for } i \in \{1, 2\}.$$

Since  $F_1^*(w) \geq 0 = F_0(w)$  for all  $w \in J_1$ ,

$$I_1 = \int_x^y \left( \frac{w-x}{z-x} \right) dw = \frac{1}{2} \frac{(y-x)^2}{z-x} = \frac{(y-x)^2(z-y)^2}{2(z-x)(z-y)^2}. \quad (27)$$

The calculation of  $I_2$  depends on the relative magnitudes of  $F_1^*(y) = (y-x)/(z-x)$  and  $F_0(y) = z/(z-y)$ . We note that

$$F_1^*(y) \leq F_0(y) \quad \text{if and only if} \quad y(x+z) \leq y^2 + z^2. \quad (28)$$

When  $F_1^*(y) \leq F_0(y)$  the quantities  $a = \frac{z}{z-y} - \frac{y-x}{z-x}$ ,  $b = -\frac{y}{z-y}$  and  $l = z-y$  are all nonnegative, so applying applying (9) after the change of variable  $u = w - y$  yields

$$\begin{aligned} I_2 &= \int_y^z \left| \frac{w-x}{z-x} - \frac{z}{z-y} \right| dw = \left( \frac{z-y}{2} \right) \frac{\left( \frac{z}{z-y} - \frac{y-x}{z-x} \right)^2 + \left( \frac{y}{z-y} \right)^2}{1 - \frac{y-x}{z-x}} \\ &= \frac{1}{2}(z-x) \left( \left( \frac{z}{z-y} - \frac{y-x}{z-x} \right)^2 + \left( \frac{y}{z-y} \right)^2 \right) \\ &= \frac{(z(z-x) - (y-x)(z-y))^2 + (y(z-x))^2}{2(z-x)(z-y)^2} \\ &= \frac{(y^2 + z^2)(z-x)^2 - 2z(z-x)(y-x)(z-y) + (y-x)^2(z-y)^2}{2(z-x)(z-y)^2}. \end{aligned} \quad (29)$$

Adding (27) to (29) yields

$$\begin{aligned} \|m_1^* - m_0\|_1 &= \frac{(y^2 + z^2)(z-x)^2 - 2z(z-x)(y-x)(z-y) + 2(y-x)^2(z-y)^2}{2(z-x)(z-y)^2} \\ &= \frac{(z^4 - 2yz^3 + x^2z^2 - 2x^2yz) + 5y^2z^2 + 3x^2y^2 - 4xy^3 + 4xy^2z - 4xyz^2 + 2y^4 - 4y^3z}{2(z-x)(z-y)^2} \end{aligned}$$

and now, subtracting from (26) and simplifying by noting that the terms in the parenthesis in the numerators of these two expressions are equal, we find

$$\begin{aligned} \|m_1^* - m_1\|_1 - \|m_1^* - m_0\|_1 &= \frac{-4y^2z^2 - 2x^2y^2 + 4xy^3 - 4xy^2z + 4xyz^2 - 2y^4 + 4y^3z}{2(z-x)(z-y)^2} \\ &= \frac{-y(y-x)(y^2 + 2z^2 - y(x+2z))}{(z-x)(z-y)^2}. \end{aligned} \quad (30)$$

The denominator in (30) is positive, as is  $-y$  and  $y-x$ . For the remaining term (28) yields

$$y^2 + 2z^2 - y(x+2z) \geq z(z-y) > 0;$$

hence (30) is positive, thus proving (25) when  $F_1^*(y) \leq F_0(y)$ .

When  $F_1^*(y) > F_0(y)$  and therefore  $y^2 + z^2 < y(x+z)$  by (28), we have  $F_1^*(w) \geq F_0(w)$  for all  $w \in [x, z]$  and hence

$$\begin{aligned} \|m_1^* - m_0\|_1 &= \int_x^z |F_1^*(w) - F_0(w)| dw = \int_x^z (F_1^*(w) - F_0(w)) dw = \int_x^z F_1^*(w) dw - \int_x^z F_0(w) dw \\ &= \int_x^z \frac{w-x}{z-x} dw - \int_y^z \frac{z}{z-y} dw = \frac{1}{2} \frac{(z-x)^2}{z-x} - z = \frac{1}{2}(z-x) - z \\ &= -\frac{x+z}{2}. \end{aligned} \quad (31)$$

Now, since  $z^2 \geq 0$

$$(x+z)(x-z) = x^2 - z^2 \leq z^2 + x^2,$$

noting that  $z - x > 0$ , dividing by  $2(z - x)$  yields under the case at hand, by (31) and Lemma 1.2, that

$$\|m_1^* - m_0\|_1 = -\frac{x + z}{2} \leq \frac{z^2 + x^2}{2(z - x)} = \|m_0^* - m_0\|_1,$$

thus proving inequality (25) when  $F_1^*(y) > F_0(y)$ , and therefore the lemma.  $\blacksquare$

### Theorem 3.1

$$B(D_3) = 1.$$

**Proof:** Let  $X \in D_3$  be arbitrary and suppose  $X$  is supported on the three points  $x < y < z$ . Lemma 1.2 shows that  $B(X) = 1$  if  $X$  is supported on two points, so we may assume that  $X$  gives positive probability to  $x, y$  and  $z$ . We first prove

$$B(X) \leq 1 \quad \text{when } X \in D_3 \text{ is positively supported on the nonzero points } x, y, z. \quad (32)$$

That  $EX = 0$  implies  $x < 0 < z$ . After proving (32) we handle the remaining case where  $y = 0$  by a continuity argument.

Let  $X$  be supported on  $x < y < z$  with  $y \neq 0$ . Lemma 1.3 with  $a = -1$  implies  $B(-X) = B(X)$ , so we may assume without loss of generality that  $x < y < 0 < z$ . Let  $m_1$  and  $m_0$  be the unique mean zero distributions supported on  $\{x, z\}$  and  $\{y, z\}$ , respectively, and let  $\mathcal{L}(X_1) = m_1$  and  $\mathcal{L}(X_0) = m_0$ . As generally every mean zero distribution having no atom at zero can be represented as a mixture of mean zero two point distributions (as in the Skorohod representation, see [2]), letting

$$\mathcal{L}(X_\alpha) = \alpha m_1 + (1 - \alpha) m_0, \quad (33)$$

we have  $\mathcal{L}(X) = \mathcal{L}(X_\alpha)$  for some  $\alpha \in [0, 1]$ ; in fact, in this particular case one may verify directly that  $P(X = x)/P(X_1 = x) \in (0, 1)$  and that (33) holds when  $\alpha$  assumes this value. Therefore to prove (32) it suffices to show

$$B(X_\alpha) \leq 1 \quad \text{for all } \alpha \in [0, 1]. \quad (34)$$

By Lemma 1.2

$$EX_1^2 = -zx \quad \text{and} \quad EX_0^2 = -zy \quad (35)$$

and by (33) the variance of  $X_\alpha$  is given by

$$EX_\alpha^2 = \alpha EX_1^2 + (1 - \alpha) EX_0^2 = -(\alpha zx + (1 - \alpha) zy) = -z(\alpha x + (1 - \alpha)y). \quad (36)$$

Applying Theorem 2.1 with  $S = \{0, 1\}$  and  $\mu$  the probability measure putting mass  $\alpha$  and  $1 - \alpha$  on the points 1 and 0, respectively, in view of (35) and (36),  $m_\alpha^*$ , the  $X_\alpha$  zero bias distribution is given by the mixture

$$m_\alpha^* = \beta m_1^* + (1 - \beta) m_0^* \quad \text{where} \quad \beta = \frac{\alpha x}{\alpha x + (1 - \alpha)y}. \quad (37)$$

Since  $x < y < 0$  we have

$$\frac{\beta}{1 - \beta} = \frac{\alpha}{1 - \alpha} \frac{x}{y} > \frac{\alpha}{1 - \alpha} \quad \text{and therefore} \quad \beta > \alpha. \quad (38)$$

Let  $F_1, F_0, F_1^*$  and  $F_0^*$  denote the distribution functions of  $m_1, m_0, m_1^*$  and  $m_0^*$ , respectively. Let  $U$  be a standard uniform variable and, with the inverse functions below given by (16), set

$$(Y_1, Y_0, Y_1^*, Y_0^*) = (F_1^{-1}(U), F_0^{-1}(U), (F_1^*)^{-1}(U), (F_0^*)^{-1}(U)).$$

Then  $Y_i =_d X_i, Y_i^* =_d X_i^*$  for  $i \in \{1, 2\}$ , and by (18), all pairs of the variables  $Y_1, Y_0, Y_1^*, Y_0^*$  achieve the  $L^1$  distance between their respective distributions. Now, recalling (38), let  $(Y_\alpha, Y_\alpha^*)$  be defined on the same space with joint distribution given by the mixture

$$\mathcal{L}(Y_\alpha, Y_\alpha^*) = \alpha \mathcal{L}(Y_1, Y_1^*) + (1 - \beta) \mathcal{L}(Y_0, Y_0^*) + (\beta - \alpha) \mathcal{L}(Y_0, Y_1^*).$$

Then  $(Y_\alpha, Y_\alpha^*)$  has marginals  $Y_\alpha =_d X_\alpha$  and  $Y_\alpha^* =_d Y_\alpha^*$ , hence by (17)

$$\|m_\alpha^* - m_\alpha\|_1 \leq \alpha \|m_1^* - m_1\|_1 + (1 - \beta) \|m_0^* - m_0\|_1 + (\beta - \alpha) \|m_1^* - m_0\|_1. \quad (39)$$

Lemma 1.2 shows  $G(X_i) = 1$ , that is, that  $E|X_i^3| = 2EX_i^2\|m_i^* - m_i\|_1$  for  $i = 1, 2$ , so (33) gives

$$E|X_\alpha^3| = 2(\alpha EX_1^2\|m_1^* - m_1\|_1 + (1 - \alpha)EX_0^2\|m_0^* - m_0\|_1),$$

and now by (35), (36) and (37) we find

$$\frac{E|X_\alpha^3|}{2EX_\alpha^2} = \frac{\alpha x\|m_1^* - m_1\|_1 + (1 - \alpha)y\|m_0^* - m_0\|_1}{\alpha x + (1 - \alpha)y} = \beta\|m_1^* - m_1\|_1 + (1 - \beta)\|m_0^* - m_0\|_1. \quad (40)$$

Lemma 3.1 shows that the right hand side, and therefore the left hand side, of (39) is bounded by (40), that is, that  $B(X_\alpha) = 2EX_\alpha^2\|m_\alpha^* - m_\alpha\|_1/E|X_\alpha^3| \leq 1$ , completing the proof of (34), and hence of (32).

Lastly we consider the case where the mean zero random variable  $X$  is positively supported on  $\{x, 0, z\}$  with  $x < 0 < z$  and  $P(X = 0) = q \in (0, 1)$ . For  $n \in \mathbb{N}$  let

$$Y_n = X\mathbf{1}(X \neq 0) + n^{-1}\mathbf{1}(X = 0) \quad \text{and} \quad X_n = Y_n - EY_n.$$

As  $n \rightarrow \infty$  we see that  $Y_n \rightarrow_{a.s.} X$  and  $EY_n = q/n \rightarrow 0$  so that  $X_n \rightarrow_{a.s.} X$ , and the bounded convergence theorem shows that  $\{X_n\}_{n \in \mathbb{N}}$  satisfies the hypothesis of Lemma 2.2. Hence  $B(X_n) \rightarrow B(X)$  as  $n \rightarrow \infty$ . For all  $n \in \mathbb{N}$  such that  $1/n < z$  the distribution of  $X_n$  is positively supported on the three distinct, nonzero points  $x - q/n < (1 - q)/n < z - q/n$ , so by (32)  $B(X_n) \leq 1$  for all such  $n$ . Therefore the limit  $B(X)$  is also bounded by 1. ■

## 4 Lower Bound

By (1) with  $m = 1$  and  $\mathcal{L}(X) = G \in \mathcal{F}_1$ ,

$$\|F_n - \Phi\|_1 \leq \frac{c_1 E|X^3|}{\sqrt{n}} \quad \text{for all } n \in \mathbb{N},$$

and in particular for  $n = 1$

$$c_1 \geq \frac{\|F_1 - \Phi\|_1}{E|X^3|} = \frac{\|G - \Phi\|_1}{E|X^3|}. \quad (41)$$

Motivated by Theorem 2.3, that two point distributions achieve the suprema of  $B(G)$ , for  $p \in (0, 1)$  let

$$X = \frac{B - p}{\sqrt{pq}},$$

where  $B$  is a Bernoulli variable with  $P(B = 1) = p = 1 - P(B = 0)$ . The distribution function  $G_p$  of  $X$  is given by

$$G_p(x) = \begin{cases} 0 & \text{for } x \leq -\sqrt{\frac{p}{q}} \\ q & \text{for } -\sqrt{\frac{p}{q}} < x \leq \sqrt{\frac{q}{p}} \\ 1 & \text{for } \sqrt{\frac{q}{p}} < x \end{cases}$$

and therefore the  $L^1$  distance between  $G_p$  and the standard normal is given by

$$\|G_p - \Phi\|_1 = \int_{-\infty}^{-\sqrt{\frac{p}{q}}} \Phi(x) dx + \int_{-\sqrt{\frac{p}{q}}}^{\sqrt{\frac{q}{p}}} |\Phi(x) - q| dx + \int_{\sqrt{\frac{q}{p}}}^{\infty} |\Phi(x) - 1| dx.$$

As  $G_p \in \mathcal{F}_1$  for all  $p \in (0, 1)$  and  $E|X^3| = (p^2 + q^2)/\sqrt{pq}$ , letting

$$\psi(p) = \frac{\sqrt{pq}}{p^2 + q^2} \|G_p - \Phi\|_1 \quad \text{for } p \in (0, 1)$$

inequality (41) gives  $c_1 \geq \psi(p)$  for all  $p \in (0, 1)$ , and  $\psi(1/2)$  yields (2).

## 5 Acknowledgment

The author would like to sincerely thank Sergey Utev for helpful suggestions regarding the paper.

## Bibliography

1. Dedecker, J., and Rio, E. (2008) On mean central limit theorems for stationary sequences *Inst. H. Poincar Probab. Statist.*, **44**, pp. 693-726.
2. Durrett, R. (2005) Probability: Theory and Examples, 3<sup>rd</sup> edition, Brooks Cole.
3. Esseen, C. G. (1958) On mean central limit theorems. *Kungl. Tekn. Hgsk. Handl. Stockholm*, **121**, pp. 1-30.
4. Erickson, R. V. (1974)  $L_1$  bounds for asymptotic normality of  $m$ -dependent sums using Stein's technique. *Ann. Probab.* **2**, pp. 522-529.
5. Goldstein, L. (2004) Normal approximation for hierarchical sequences. *Annals of Applied Probability*, **14**, pp. 1950-1969. arXiv:math/0503549v1 [math.PR]
6. Goldstein, L. (2005) Berry Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *Jour. Appl. Probab.*, **42**, pp. 661-683. arXiv:math/0511510v1 [math.PR]

7. Goldstein, L. (2007)  $L^1$  bounds in normal approximation. *Ann. Probab.* **35**, pp. 1888–1930. arXiv:0710.3262v1 [math.PR]
8. Goldstein, L. (2009) A probabilistic proof of the Lindeberg-Feller central limit theorem *Amer. Math. Monthly*, to appear.
9. Goldstein, L., and Reinert, G. (1997) Stein’s method and the zero bias transformation with application to simple random sampling *Ann. Appl. Probab.*, **7**, 935-952. arXiv:math/0510619v1 [math.PR]
10. Hoeffding, W. (1955). The extrema of the expected value of a function of independent random variables. *Ann. Math. Stat.*, **26**, pp. 268-275.
11. Ho, S. T. and Chen, Louis H. Y. (1978). An  $L_p$  bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.* **6**, pp. 231-249.
12. Ibragimov, I. and Linnik, I. (1971). Independent and stationary sequences of random variables. Wolters-Noordhoff.
13. Lefvre, C. and Utev, S. (2003). Exact norms of a Stein-type operator and associated stochastic orderings. *Probab. Theory Related Fields*, **127**, pp. 353-366.
14. Rachev, S. T. (1984) The Monge-Kantorovich transference problem and its stochastic applications. *Theory Probab. Appl.* **29**, pp. 647-676.
15. Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Proc. Sixth Berkeley Symp. Math. Statist. Probab. **2** 583-602, Univ. California Press, Berkeley.
16. Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, pp. 1135-1151.
17. Stein, C. (1986). Approximate Computation of Expectations. IMS, Hayward, CA.
18. Zolotarev, V. M. (1964) On Asymptotically Best Constants in Refinements of Mean Limit Theorems. *Theory Probab. Appl.*, **9**, pp. 268-276.